

Probability Theory and Likelihood Ratios

Simone Gittelsohn, Ph.D., simone.gittelsohn@nist.gov

Michael Coble, Ph.D., michael.coble@nist.gov

Acknowledgement

I thank Michael Coble, Bruce Weir and John Buckleton for their helpful discussions.

Disclaimer

Points of view in this presentation are mine and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.

Probability Theory

Vocabulary and Notation



rolling a 6 and a 1

it will snow tomorrow



observing a peak for allele 8

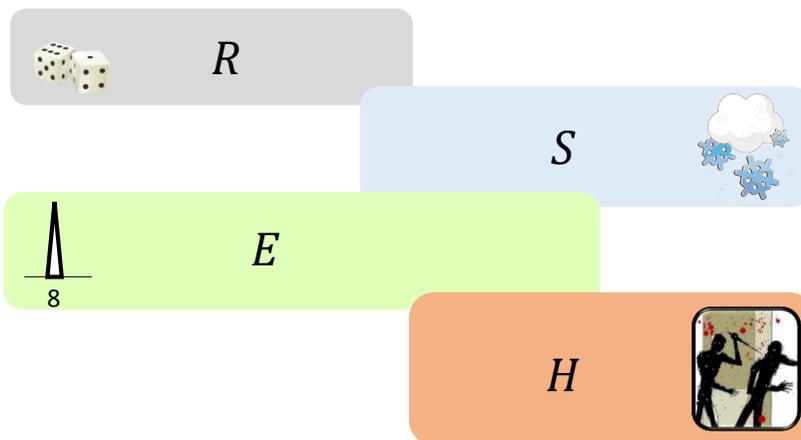
Mr. X stabbed Mr. Y



Vocabulary and Notation

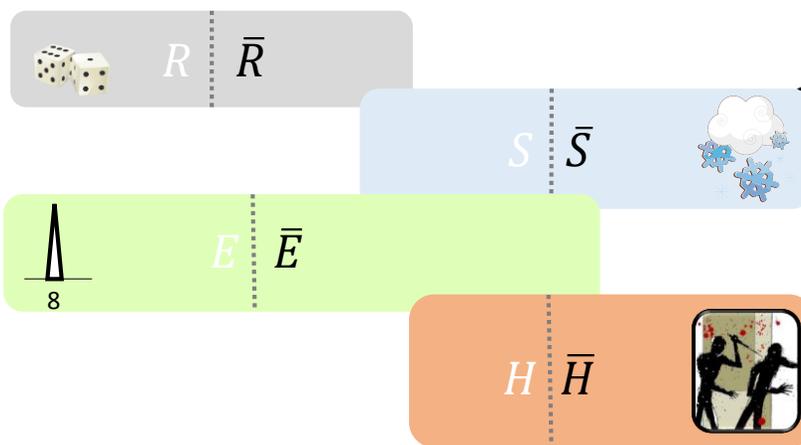
EVENT

PROPOSITION



Vocabulary and Notation

COMPLEMENT event or proposition



Vocabulary and Notation

PROBABILITY



$\Pr(R)$

$\Pr(S)$



$\Pr(E)$

$\Pr(H)$



Laws of Probability

Law #1: A probability can take any value between 0 and 1, including 0 and 1.

certain



1 certainty that statement is true

0.75

0.66

0.5

0.33

0.25

0

certainty that statement is false

impossible

EXAMPLE: rolling a 6-sided die

$\Pr(1,2,3,4,5 \text{ or } 6) = 1$

$\Pr(7) = 0$

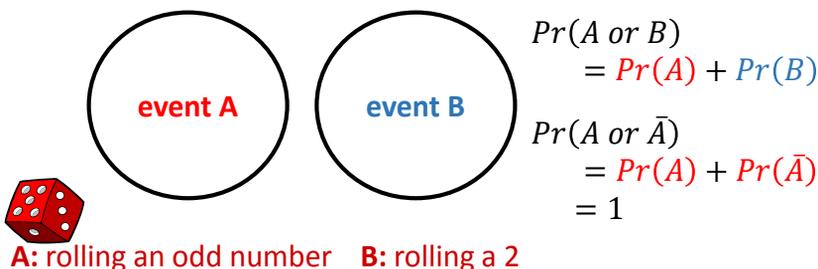


Laws of Probability

Or

Law #2: The probability of event A **or** event B occurring is equal to the probability of event A **plus** the probability of event B minus the probability of event A and B.

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$



Second Law of Probability

Blood groups:

O	A	B	AB
0.45	0.40	0.11	0.04

1. What is the probability that a person drawn randomly from this population is blood group type A or AB?
2. What is the probability that a person drawn randomly from this population is not blood group type AB?

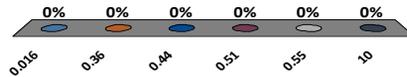
data from: <http://www.redcrossblood.org/learn-about-blood/blood-types>

1. What is the probability that a person drawn randomly from this population is blood group type A or AB?

Blood groups:

O	A	B	AB
0.45	0.40	0.11	0.04

- A. 0.016
- B. 0.36
- C. 0.44
- D. 0.51
- E. 10

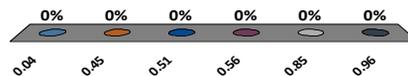


2. What is the probability that a person drawn randomly from this population is not blood group type AB?

Blood groups:

O	A	B	AB
0.45	0.40	0.11	0.04

- A. 0.04
- B. 0.45
- C. 0.51
- D. 0.85
- E. 0.96

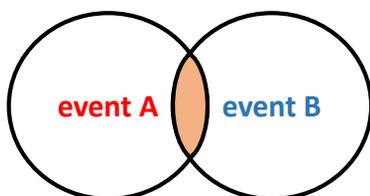


Laws of Probability

Or

Law #2: The probability of event A **or** event B occurring is equal to the probability of event A **plus** the probability of event B minus the probability of event A and B.

$$Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A \text{ and } B)$$



$$Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A \text{ and } B)$$



A: rolling an odd number

B: rolling a number greater than 2

Second Law of Probability

Blood groups:

O	A	B	AB
0.45	0.40	0.11	0.04

- What is the probability that a person drawn randomly from this population has a blood group that contains an A (groups A and AB) or a blood group that contains a B (groups B and AB)?

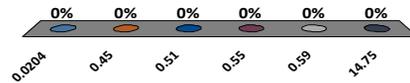
data from: <http://www.redcrossblood.org/learn-about-blood/blood-types>

3. What is the probability that a person drawn randomly from this population has a blood group that contains an A (groups A and AB) or a blood group that contains a B (groups B and AB)?

Blood groups:

O	A	B	AB
0.45	0.40	0.11	0.04

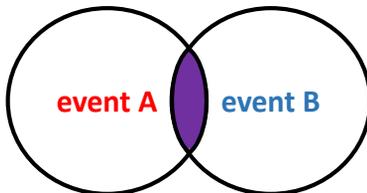
- A. 0.0204
- B. 0.45
- C. 0.55
- D. 0.59
- E. 14.75



Laws of Probability *and*

Law #3: The probability of event A *and* event B occurring is equal to the probability of event A *times* the probability of event B.

$$\begin{aligned} \Pr(A \text{ and } B) &= \Pr(A) \times \Pr(B|A) \\ &= \Pr(B) \times \Pr(A|B) \end{aligned}$$

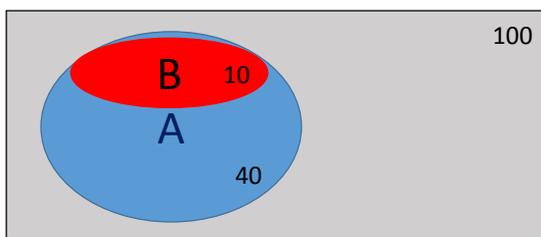


A: rolling an odd number

B: rolling a 3 or less



Conditional Probabilities



What is $Pr(B)$? $\frac{10}{100}$

What is $Pr(B|A)$? $\frac{10}{40}$

Third Law of Probability (independent events)

Blood groups:

O	A	B	AB
0.45	0.40	0.11	0.04

Rh factor:

+	-
0.82	0.18

4. What is the probability that a person drawn randomly from this population is blood group type A and Rh+?

data from: <http://www.redcrossblood.org/learn-about-blood/blood-types>

4. What is the probability that a person drawn randomly from this population is blood group type A and Rh+?

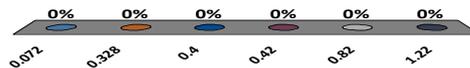
Blood groups:

O	A	B	AB
0.45	0.40	0.11	0.04

Rh factor:

+	-
0.82	0.18

- A. 0.328
- B. 0.4
- C. 0.42
- D. 0.82
- E. 1.22



Third Law of Probability

hair color:

black	brown	red	blond
0.16	0.46	0.12	0.26

eye color:

brown	blue	hazel	green
0.39	0.36	0.15	0.10

5. What is the probability that a person drawn randomly from this population has blond hair and blue eyes?

data from: <http://data-sorcery.org/2009/06/14/chi-square-test/>

Third Law of Probability (dependent events)

hair color:

black	brown	red	blond
0.16	0.46	0.12	0.26

eye color of blond haired people:

brown	blue	hazel	green
0.05	0.79	0.06	0.10

5. What is the probability that a person drawn randomly from this population has blond hair and blue eyes?

data from: <http://data-sorcery.org/2009/06/14/chi-square-test/>

5. What is the probability that a person drawn randomly from this population has blond hair and blue eyes?

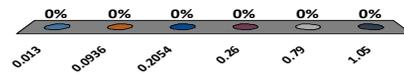
hair color:

black	brown	red	blond
0.16	0.46	0.12	0.26

eye color of blond haired people:

brown	blue	hazel	green
0.05	0.79	0.06	0.10

- A. 0.013
- B. 0.0936
- C. 0.2054
- D. 0.26
- E. 1.05



Hardy-Weinberg Law

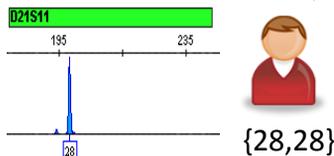
Homozygote

$$Pr(28,28) = Pr(\text{paternal allele} = 28) \times Pr(\text{maternal allele} = 28)$$



$$= p_{28} \times p_{28}$$

$$= p_{28}^2$$



$$p_{28} = 0.159$$

$$Pr(28,28) = (0.159)^2 = 0.025$$

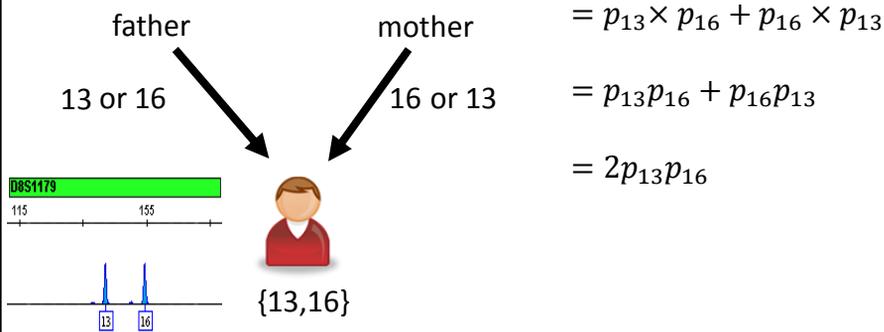
Hardy-Weinberg Law

Heterozygote

$$Pr(13,16)$$

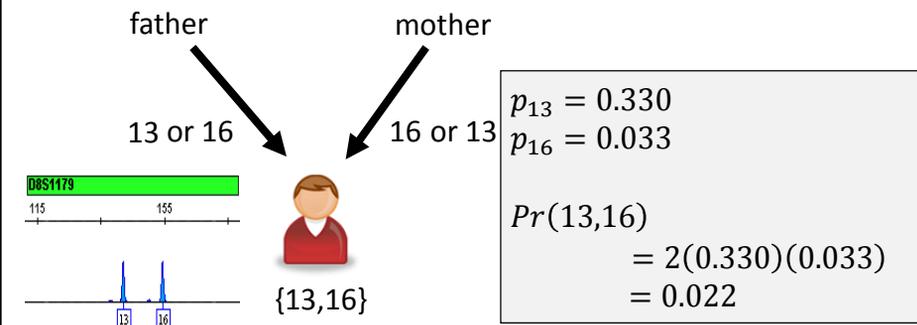
$$= Pr(\text{paternal allele} = 13) \times Pr(\text{maternal allele} = 16)$$

$$\text{or} + Pr(\text{paternal allele} = 16) \times Pr(\text{maternal allele} = 13)$$



Hardy-Weinberg Law

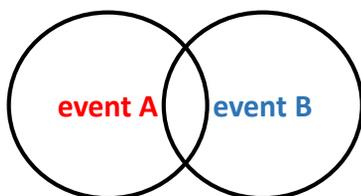
Heterozygote



Laws of Probability

Law of total probability or the extension of the conversation:

The probability of event A can be partitioned into the sum of probabilities of event A conditioned on mutually exclusive and exhaustive events.



$$\Pr(A) = \Pr(A|B) \times \Pr(B) + \Pr(A|\bar{B}) \times \Pr(\bar{B})$$

Law of Total Probability

hair color:

black	brown	red	blond
0.16	0.46	0.12	0.26

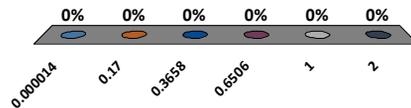
eye color:

	brown	blue	hazel	green
black hair	0.69	0.17	0.10	0.04
brown hair	0.46	0.24	0.20	0.10
red hair	0.43	0.19	0.19	0.19
blond hair	0.05	0.79	0.06	0.10

6. What is the probability that a person drawn randomly from this population has blue eyes?

6. What is the probability that a person drawn randomly from this population has blue eyes?

- A. 0.000014
- B. 0.17
- C. 0.3658
- D. 0.6506
- E. 1



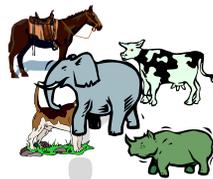
Conditional Probability

H : the animal is an elephant

E : the animal has four legs

- 1 leg
- 2 legs
- 3 legs
- 4 legs

given:



given:
4 legs

$$\Pr(E|H)$$

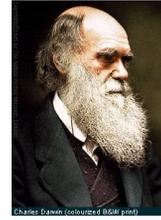
$$\sim 1$$

\neq

$$\Pr(H|E)$$

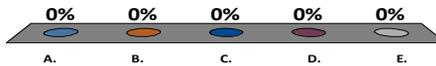
$$\frac{1}{5,000}$$

H : This man is Santa Claus.
 E : This man has a white beard.

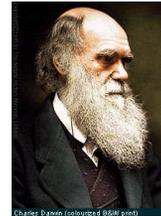


$\Pr(E|H) = ?$

- A. The probability that this man is Santa Claus given that he has a white beard.
- B. The probability that this man has a white beard given that he is Santa Claus.
- C. both A) and B) are correct
- D. none of the above
- E. I am completely lost

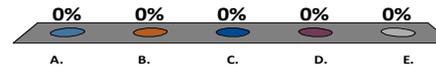


H : This man is Santa Claus.
 E : This man has a white beard.



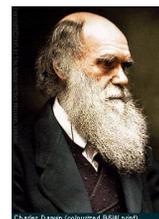
$\Pr(H|E) = ?$

- A. The probability that this man is Santa Claus given that he has a white beard.
- B. The probability that this man has a white beard given that he is Santa Claus.
- C. both A) and B) are correct
- D. none of the above
- E. I am completely lost

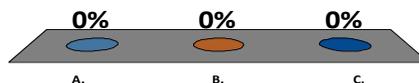


H: This man is Santa Claus.
E: This man has a white beard.

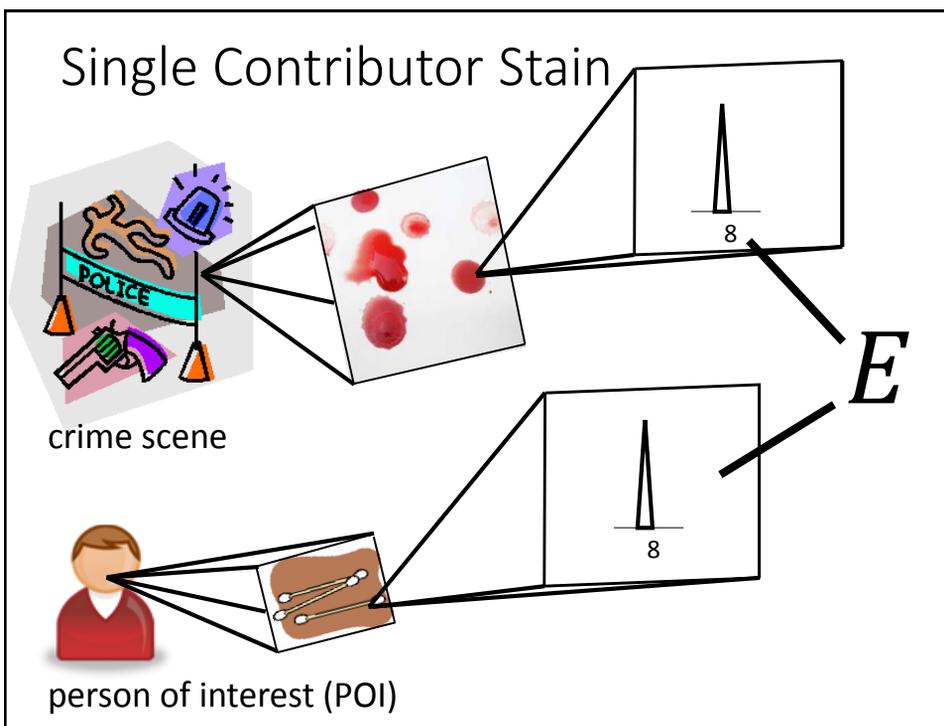
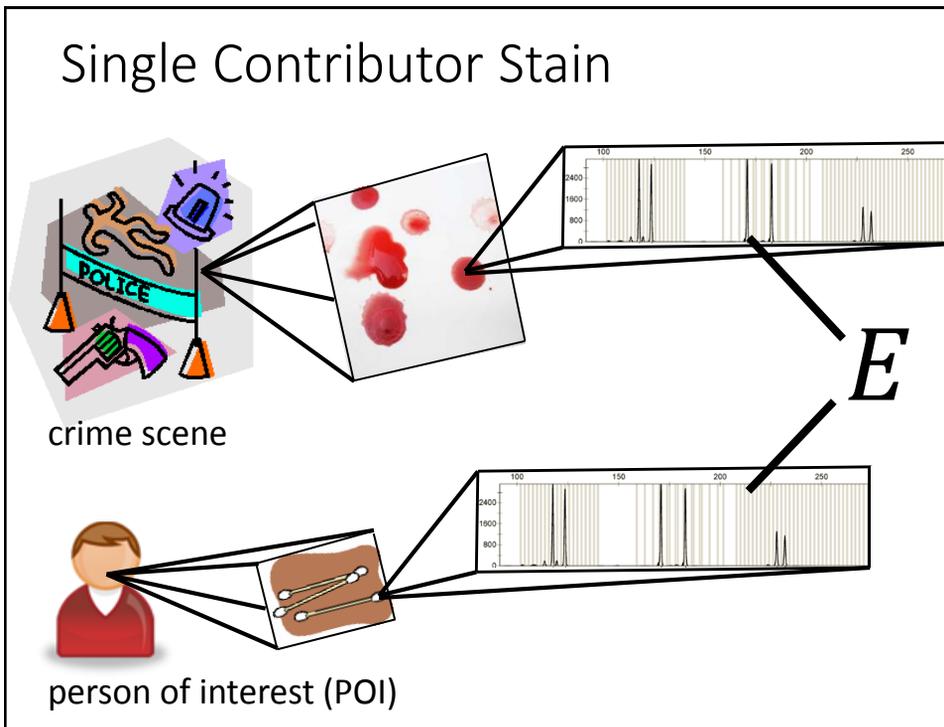
Is $\Pr(E|H) = \Pr(H|E)$?



- A. Yes
- B. No
- C. I am completely lost



Likelihood Ratio



There are two sides to every story...



H_p : The crime stain came from the person of interest (POI).



H_d : The crime stain did not come from the POI. It came from some other person.

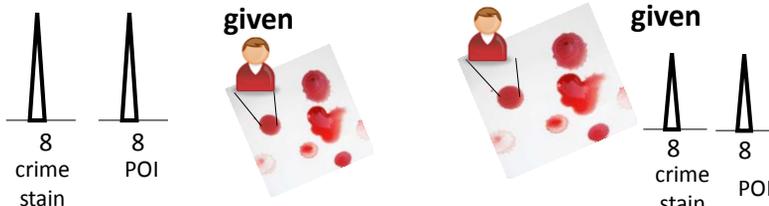
prosecution's proposition

defense's proposition

Conditional Probability

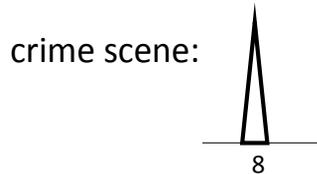
H_p : The crime stain came from the POI.

E : The DNA typing results of the crime stain and the POI's sample both show a peak for allele 8.



$$\Pr(E|H_p) \neq \Pr(H_p|E)$$

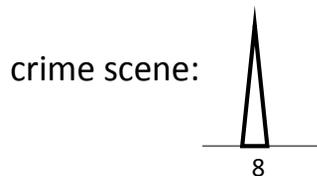
We have the DNA typing results of only one marker. This marker is called *NEW*, and we don't know anything about the alleles at this locus.



Does the DNA on the crime scene come from Albert?

- A) Yes
- B) No
- C) I need more information
- D) I am completely lost

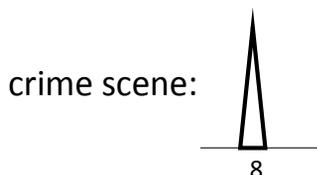
We have the DNA typing results of only one marker. This marker is called *NEW*.



If Albert left this DNA on the crime scene, we would expect to observe a peak for allele 8.

The probability of this observation if Albert left this DNA on the crime scene is 1.

We have the DNA typing results of only one marker.
This marker is called *NEW*.

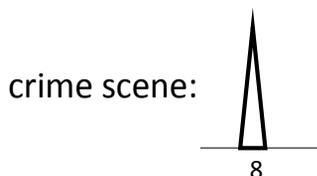


In the population of potential donors, 1 out of 10 people have genotype {8,8}.



Observation supports the proposition that Albert is the donor.

We have the DNA typing results of only one marker.
This marker is called *NEW*.



In the population of potential donors, everyone has genotype {8,8}.



Observation provides no information on the donor of the crime stain.

Likelihood Ratio (LR)

given or if
↓
 $\Pr(E|H_p)$
 $\Pr(E|H_d)$

the probability of observing the
DNA typing results if the
prosecution's proposition is true

divided by

the probability of observing the
DNA typing results if the defense's
proposition is true.

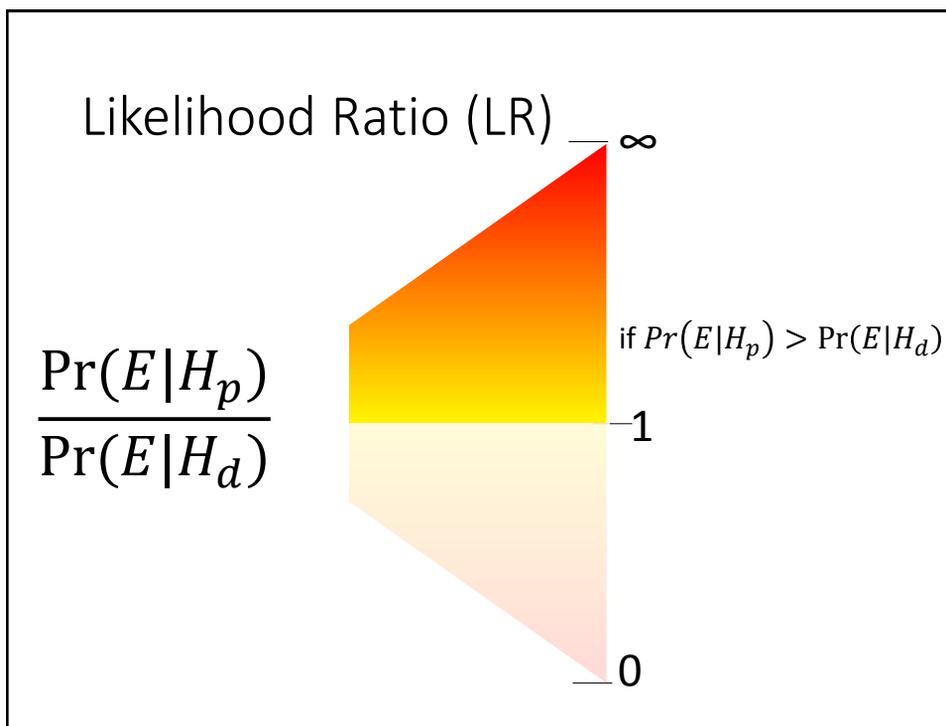
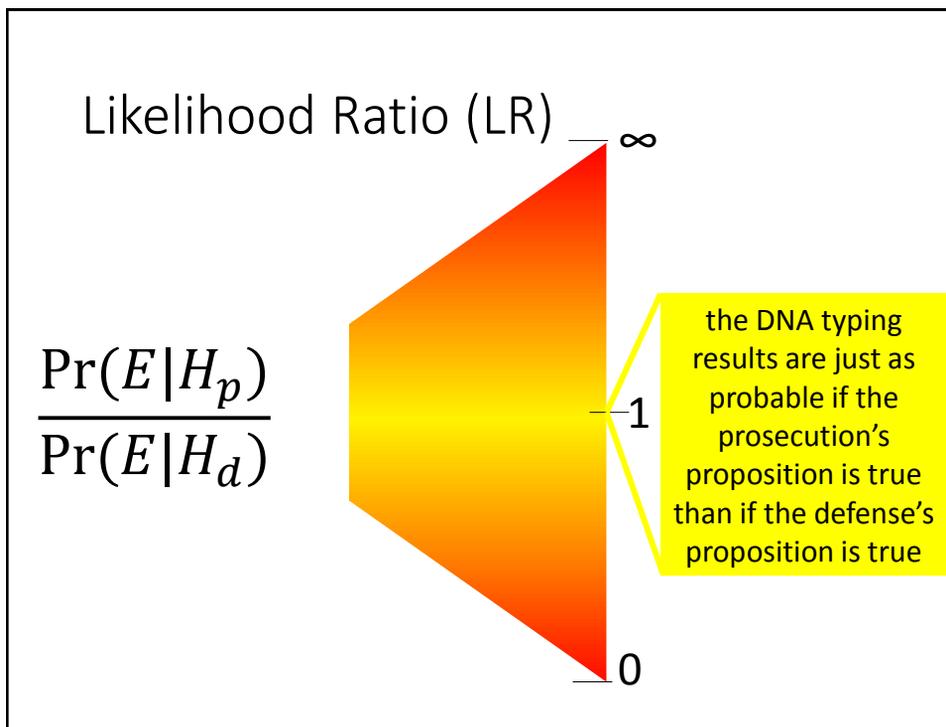
Likelihood Ratio (LR)

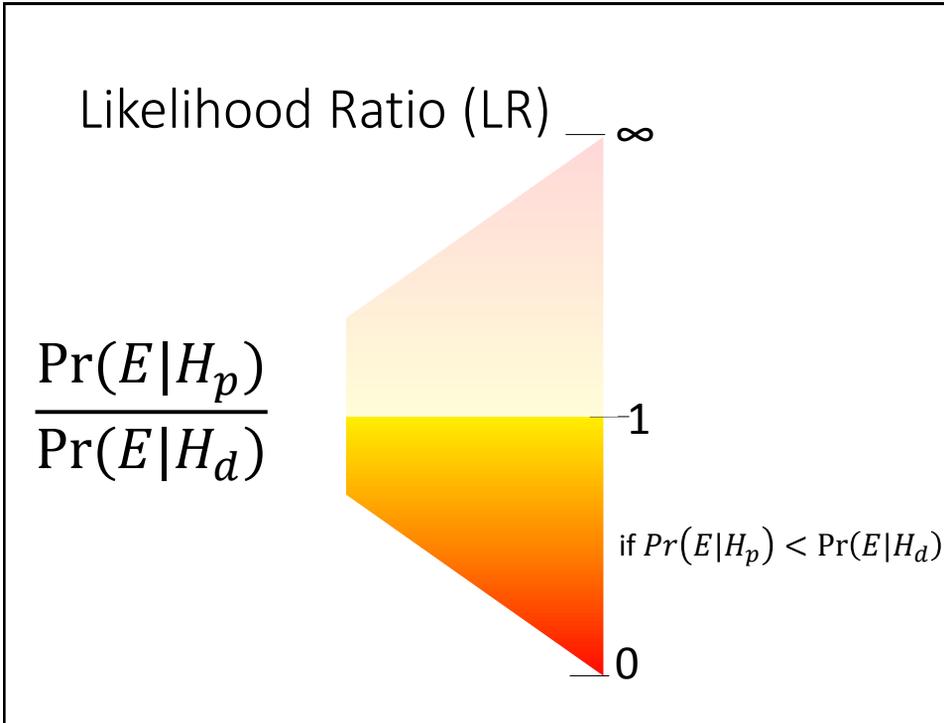
$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

The probability of observing the
DNA typing results given that the
prosecution's proposition is true

divided by

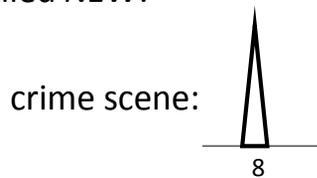
the probability of observing the
DNA typing results given that the
defense's proposition is true.





Likelihood Ratio (LR)

We have the DNA typing results of only one marker, called *NEW*.



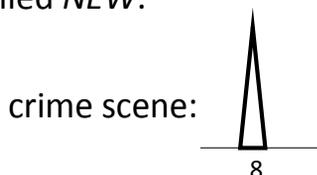
If Albert left this DNA on the crime scene, we would expect to observe a peak for allele 8.

The probability of this observation if Albert left this DNA on the crime scene is 1.

$$\Pr(E|H_p) = 1$$

Likelihood Ratio (LR)

We have the DNA typing results of only one marker, called *NEW*.



In the population of potential donors, 1 out of 10 people have genotype {8,8}.

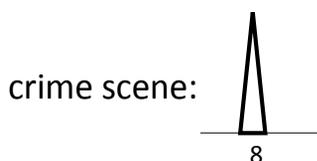
$$\Pr(E|H_p) = 1$$

$$\Pr(E|H_d) = \frac{1}{10}$$

$$LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} = \frac{1}{\frac{1}{10}} = 10$$

Likelihood Ratio (LR)

We have the DNA typing results of only one marker, called *NEW*.



In the population of potential donors, everyone has genotype {8,8}.

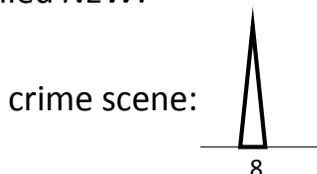
$$\Pr(E|H_p) = 1$$

$$\Pr(E|H_d) = 1$$

$$LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} = \frac{1}{1} = 1$$

Likelihood Ratio (LR)

We have the DNA typing results of only one marker, called *NEW*.



In the population of potential donors, Albert is the only one who has genotype {8,8}.

$$\Pr(E|H_p) = 1$$

$$\Pr(E|H_d) = 0$$

$$LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} = \frac{1}{0} = \infty \text{ infinity}$$

individualization

Likelihood Ratio (LR)

We have the DNA typing results of only one marker, called *NEW*.



In the population of potential donors, 1 out of 10 people have genotype {8,8}.

$$\Pr(E|H_p) = 0$$

$$\Pr(E|H_d) = \frac{1}{10}$$

$$LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} = \frac{0}{\frac{1}{10}} = 0$$

exclusion

True or False?

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)} = 100, \text{ or abbreviated as } LR = 100$$

A)

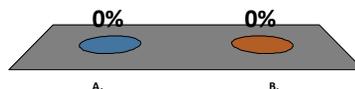
Given the available information, the probability of these DNA typing results is 100 times greater if the prosecution's proposition is true than if the defense's proposition is true.

B)

Given the available information, these DNA typing results indicate that the probability of the prosecution's proposition being true is 100 times greater than the probability of the defense's proposition being true.

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)} = 100, \text{ or abbreviated as } LR = 100$$

- A. Given the available information, the probability of these DNA typing results is 100 times greater if the prosecution's proposition is true than if the defense's proposition is true.
- B. Given the available information, these DNA typing results indicate that the probability of the prosecution's proposition being true is 100 times greater than the probability of the defense's proposition being true.



True or False?

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)} = 100, \text{ or abbreviated as } LR = 100$$



Given the available information, the probability of these DNA typing results is 100 times greater if the prosecution's proposition is true than if the defense's proposition is true.



Given the available information, these DNA typing results indicate that the probability of the prosecution's proposition being true is 100 times greater than the probability of the defense's proposition being true.

Exercise 1: Likelihood Ratio

The DNA typing results of a biological stain recovered on a crime scene in Washington DC show the same genotype as that of a person of interest (POI) living in Washington DC. If the biological stain came from the POI, the forensic scientist would expect to obtain these typing results. According to population data, we would expect to see this genotype in one person out of 500,000. The forensic scientist formulates:

H_p : The DNA recovered on the crime scene came from the POI.

H_d : The DNA recovered on the crime scene came from someone else, unrelated to the POI.

What is the likelihood ratio for these DNA typing results with regard to propositions H_p and H_d ?

Exercise 1: Likelihood Ratio

Write a sentence describing your likelihood ratio in words:

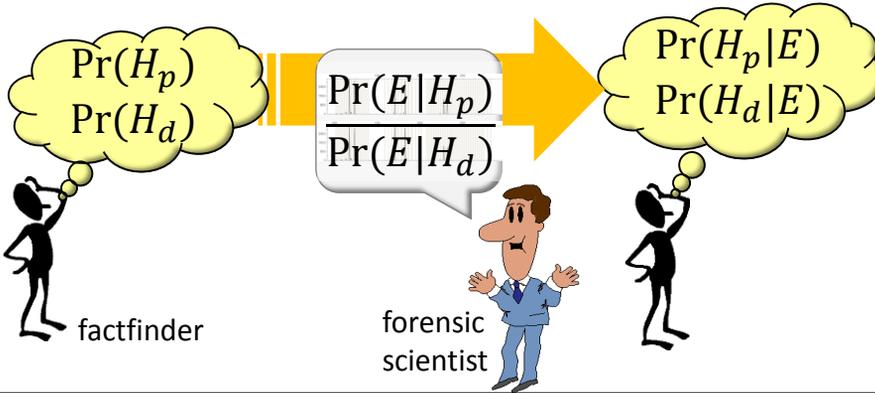
Bayes' theorem

Logical Framework for Updating Uncertainty

H_p : The crime stain came from the POI.

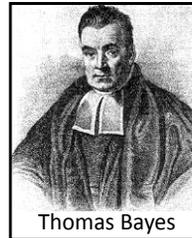
H_d : The crime stain did not come from the POI. It came from some other person.

E : The DNA typing results of the crime stain and the POI's sample both show a peak for allele 8.



Logical Framework for Updating Uncertainty

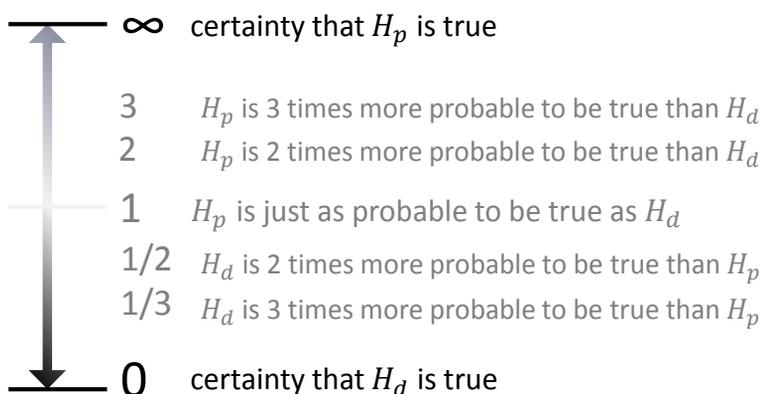
Odds form of **Bayes' theorem**:



$$\underbrace{\frac{\Pr(H_p)}{\Pr(H_d)}}_{\text{prior odds}} \times \underbrace{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}}_{\text{Likelihood Ratio}} = \underbrace{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}}_{\text{posterior odds}}$$

Logical Framework for Updating Uncertainty

odds $\frac{\Pr(H_p)}{\Pr(H_d)}$:



From Probabilities to Odds and back again

$$odds = \frac{\Pr(H)}{\Pr(not H)} = \frac{\Pr(H)}{1 - \Pr(H)}$$

$$\Pr(H) = \frac{odds}{1 + odds}$$

Transposed Conditional

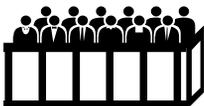
$$\frac{\Pr(H_p)}{\Pr(H_d)} \times \underbrace{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}}_{100} = \underbrace{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}}_{100}$$

The probability of these DNA typing results is 100 times greater if the prosecution's proposition is true than if the defense's proposition is true.

These DNA typing results indicate that the probability of the prosecution's proposition being true is 100 times greater than the probability of the defense's proposition being true.

Logical Framework for Updating Uncertainty

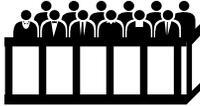
$$\underbrace{\frac{\Pr(H_p)}{\Pr(H_d)}}_{\text{prior odds}} \times \underbrace{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}}_{\text{Likelihood Ratio}} = \underbrace{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}}_{\text{posterior odds}}$$



factfinder

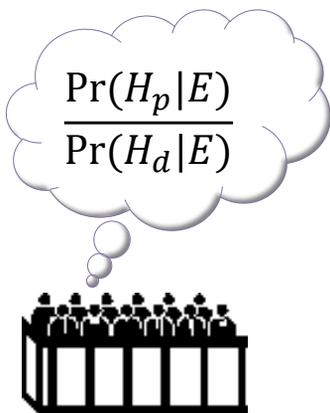


expert witness



factfinder

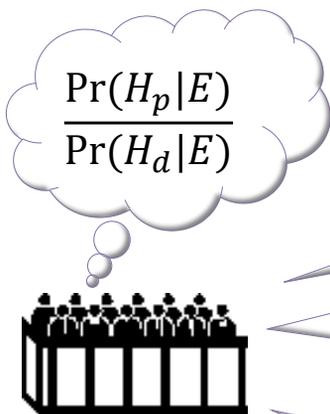
Role of the Factfinder



Given the evidence, what is the probability that the prosecution's proposition is true?

Given the evidence, what is the probability that the defense's proposition is true?

Role of the Factfinder



The chance that the crime stain came from the POI is 0.9.

The probability that the crime stain came from someone other than the POI is 0.1.

There is a 1 in 10 chance that the crime stain did not come from the POI.

Role of the Forensic Scientist

What is the probability of the analytical results if the prosecution's proposition is true?

What is the probability of the analytical results if the defense's proposition is true?

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$



Role of the Forensic Scientist

The probability of obtaining these DNA results if the crime stain came from the suspect is very close to 1.

The chance of obtaining these DNA results if the crime stain came from some other person, unrelated to the suspect, is 1 in 1 million.

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$



Prosecutor's Fallacy

$$\Pr(E|H_d, I) = 1 \text{ in } 7 \text{ million}$$



In layman's terms, just so I get this right, are you saying that the probability that the DNA that was found in the question samples came from anyone else besides A.L. is one in 7 million (...)?



State v. Lee, No. 90CA004741 (Ohio App. Dec. 5, 1990), transcript at 464

Prosecutor's Fallacy

$$\Pr(E|H_d, I) = 1 \text{ in } 10 \text{ million}$$



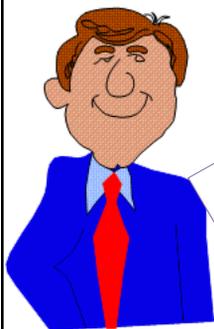
The witness concludes that the genetic profile of the two analyzed samples match perfectly, and he deduces that the probability of someone other than the suspect being the source of the trace found on the victim's cloths is 1 in 10 million.



modified from:
State of Arizona v. Michael Steven Gallegos [178 Ariz. 1; 870 P.2d 1097 (1994)]

Prosecutor's Fallacy

LR = 4.73 quadrillion



The DNA mixture profile obtained from *[the item of evidence]* is 4.73 quadrillion times more likely to have originated from *[suspect]* and *[victim/complainant]* than from an unknown individual in the U.S. Caucasian population and *[victim/complainant]*.



Holder, E.H., Leary M.L., Laub J.H. DNA for the Defense Bar 2012.

Prosecutor's Fallacy

The fallacy is to transpose the conditional:

$$\Pr(E|H_d, I) = \Pr(H_d|E, I)$$

or

$$\frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} = \frac{\Pr(H_p|E, I)}{\Pr(H_d|E, I)}$$

Prosecutor's Fallacy

which means that a low $\Pr(E|H_d, I)$ is expressed

as a low $\Pr(H_d|E, I)$ or high $\frac{\Pr(H_p|E, I)}{\Pr(H_d|E, I)}$ when the

prior odds are not necessarily equal to 1.

Defense Attorney's Fallacy



$$\Pr(E|H_d, I) = 1 \text{ in } 1,000$$

The city where the crime occurred has a population of 200,000. In this city, this genotype would be found in 200 people. Therefore the evidence merely shows that the person of interest is one of 200 people in the city who might have left the crime stain.



modified from: W.C. Thompson and E.L. Schumann. Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and defence attorney's fallacy. *Law and Human Behaviour*, 11: 167-187, 1987.

Defense Attorney's Fallacy

The fallacy is:

1. 200 individuals in the population plus the genotyped POI is equal to 201
2. To assume that each of these 200 individuals has the **same prior probability** of being the source of the crime stain as the POI
3. To assume that the **actual number** of individuals in this city having the genotype in question is equal to the **expected number** of individuals having this genotype. The actual number could be anywhere between 1 and 200,000.

Correct

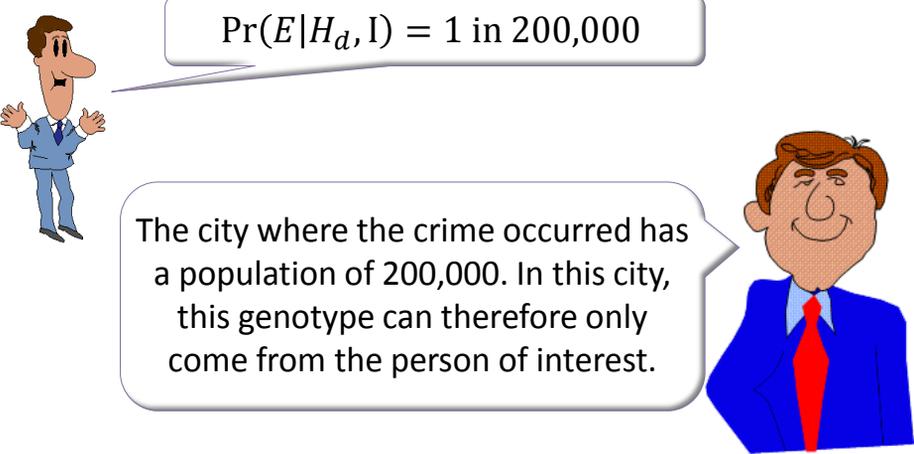


$$\Pr(E|H_d, I) = 1 \text{ in } 1,000$$

The city where the crime occurred has a population of 200,000. In this city, we would **expect** to find this genotype in 200 untyped people in addition to the POI.



Uniqueness Fallacy



$\Pr(E|H_d, I) = 1 \text{ in } 200,000$

The city where the crime occurred has a population of 200,000. In this city, this genotype can therefore only come from the person of interest.

Uniqueness Fallacy

The fallacy is:

1. 1 individual in the population plus the genotyped POI is equal to a total of two individuals
2. To assume that the **actual number** of individuals in this city having the genotype in question is equal to the **expected number** of individuals having this genotype. The actual number could be anywhere between 1 and 200,000.

Correct



$$\Pr(E|H_d, I) = 1 \text{ in } 200,000$$

The city where the crime occurred has a population of 200,000. In this city, we would **expect** to find this genotype in 1 untyped person in addition to the POI.



Exercise 1: Moot Court

- A) Correct
- B) Transposed Conditional / Prosecutor's Fallacy
- C) Defense Attorney's Fallacy
- D) Uniqueness Fallacy
- E) ???

Principles of Evidence Interpretation

1. To evaluate the uncertainty of a proposition, it is necessary to consider at least one alternative proposition.
2. Scientific interpretation is based on questions of the kind “What is the probability of the evidence given the proposition?”
3. Scientific interpretation is conditioned not only by the competing propositions, but also by the framework of circumstances within which they are to be evaluated.

I.W. Evett and B.S. Weir. *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. Sinauer Associates, Sunderland, 1998: page 29.

Conditioning on the Framework of Circumstances

$$LR = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)}$$

where I is the available information

case circumstances and evidence presented so far

- where was the crime committed
- when was the crime committed
- description of the offender(s)
- information about the commission of the crime (activities performed, modus operandi, crime scene entry/exit points)
- information about the crime scene before and after the commission of the crime
- information about how rare particular characteristics are in a population

Conditioning on the Framework of Circumstances

$$LR = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)}$$

with I :

What do we know or assume about the offender? What population does the offender come from?

What are the genetic properties of this population? What do we know about the rarity of the observed genotype in this population?



Conditioning on the Framework of Circumstances

$$LR = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)}$$

The LR will vary according to the information in I .

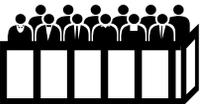
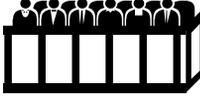


It is therefore imperative for the forensic scientist to make explicit to the court what information makes up the I in his/her LR. If the court disagrees, or new information becomes available, the forensic scientist must re-assign the probabilities forming the LR conditioned on the new I .

Logical Framework for Updating Uncertainty

Odds form of Bayes' theorem:

$$\underbrace{\frac{\Pr(H_p|I)}{\Pr(H_d|I)}}_{\text{prior odds}} = \underbrace{\frac{\Pr(E|H_p,I)}{\Pr(E|H_d,I)}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{\Pr(H_p|E,I)}{\Pr(H_d|E,I)}}_{\text{posterior odds}}$$

factfinder expert witness factfinder